Journal of Computational and Applied Mathematics 38 (1991) 185–200
North-Holland

Projected successive overrelaxation method for finite-element solutions to the Dirichlet problem for a system of nonlinear elliptic equations

Kazuo Ishihara

Department of Mathematics, Osaka Women's University, Daisen-cho, Sakai, Osaka 590, Japan

Received 30 July 1990
Revised 14 August 1990

Abstract

Ishihara, K., Projected successive overrelaxation method for finite-element solutions to the Dirichlet problem for a system of nonlinear elliptic equations, Journal of Computational and Applied Mathematics 38 (1991) 185–200.

In this paper, we consider the projected successive overrelaxation (SOR) method for obtaining finite-element solutions applied to the Dirichlet problem for a system of nonlinear elliptic equations. These equations arise in gas dynamics and chemical reactions. The Jacobian matrix for the nonlinear equations is not symmetric, so that there is no longer an associated minimization problem. A convergence proof of the projected SOR method is established by using a contraction argument, because minimization techniques are not applicable. We also discuss the optimum relaxation parameter, based upon the linear SOR theory. Finally, we show some numerical examples to indicate the effectiveness of the projected SOR method.

Keywords: System of nonlinear equations, finite-element solution, projected SOR method, convergence, optimum relaxation parameter.

1. Introduction

This paper is concerned with the projected successive overrelaxation (SOR) method for obtaining finite-element solutions applied to the Dirichlet problem for a system of nonlinear elliptic equations:

$$\begin{aligned} &\Delta u = b_1 u^{n_1} v^{n_2}, \quad &&\Delta v = b_2 u^{n_1} v^{n_2}, \quad &&\text{in } \Omega, \\ &u = g_1(x), &&v = g_2(x), &&\text{on } \Gamma. \end{aligned} \tag{1.1}$$

Here Ω is a bounded convex domain in the real n-dimensional Euclidean space $\mathbb{R}^n$, its boundary Γ is piecewise smooth, $x = (x_1, x_2, \ldots, x_n)$, Δ is the Laplace operator, b_1 and b_2 are positive constants, n_1 and n_2 are positive integers, and given functions $g_1(x)$ and $g_2(x)$ are smooth and nonnegative. These equations arise, for example, in gas dynamics and chemical reactions [1,6]. In such cases, $u(x)$ and $v(x)$ represent the chemical concentrations, so that $u(x)$ and $v(x)$ are required to be nonnegative. The uniqueness and existence of the nonnegative solution of (1.1) was established [12,15]. To avoid the trivial solution ($u(x) \equiv 0$ or $v(x) \equiv 0$), we assume that

$$\max_{x \in \Gamma} g_1(x) > 0, \qquad \max_{x \in \Gamma} g_2(x) > 0. \tag{1.2}$$

Then, by applying the maximum principles [23], the solution of (1.1) has the following constraint:

$$0 < u(x) < \max_{x \in \Gamma} g_1(x), \quad 0 < v(x) < \max_{x \in \Gamma} g_2(x), \qquad \text{in } \Omega. \tag{1.3}$$

In previous papers [7,8], we discussed monotone iterations for solving a system of nonlinear equations associated with the finite-element approximation to (1.1), based upon piecewise linear functions and piecewise constant functions. Reference [7] considered implicit monotone iterations in which a system of linear equations has to be solved at each stage. This is the disadvantage from a computational view point. Reference [8] considered explicit monotone iterations. In practical computations of a very large scale problem, it seems that the explicit monotone iterations converge too slowly. In order to overcome such faults, it is well known that SOR methods are effective for increasing the speed of convergence, and there is much literature on SOR methods. See, for instance, [5,13,14,16–22,24,25] and the references given there, which deal with systems of equations and inequalities.

By following recent papers [10,11], the objective of this paper is to present the projected SOR method, which provides a simple algorithm induced by the constraint, i.e., one need not solve a system of linear equations at each stage. In particular, its convergence proof is established by using a contraction argument, because the Jacobian matrix is not symmetric, so that there is no equivalent minimization problem unlike the recent papers [10,11] in which minimization techniques were applied. We also consider the optimum relaxation parameter, based upon the linear SOR theory. The use of the projected SOR method leads to a significant reduction of computational efforts by choosing the optimum relaxation parameter. Finally, some numerical examples are given to illustrate the validity of our results.

2. Finite-element approximation and projected SOR method

For simplicity we assume that Ω is a polyhedral domain of $\mathbb{R}^n$. As usual, we triangulate Ω in such a way that

$$\overline{\Omega} \equiv \Omega \cup \Gamma = \bigcup_{q=1}^{J} T_q,$$

where T_q, $1 \leqslant q \leqslant J$, are nondegenerate closed n-simplices whose interiors are pairwise disjoint and any one of the faces of T_q is either a face of another T_r $(r \neq q)$ or else is a portion of the boundary Γ. By P_i, $1 \leqslant i \leqslant N$ (or P_i, $N+1 \leqslant i \leqslant N+M$), we denote the vertices of a triangulation which belong to Ω (or Γ). Set

$$h_q = \text{diameter of } T_q, \quad 1 \leqslant q \leqslant J, \qquad h = \max_{1 \leqslant q \leqslant J} h_q,$$

$$\beta_q = \text{supremum of the diameter of the inscribed sphere of } T_q, \quad 1 \leqslant q \leqslant J.$$

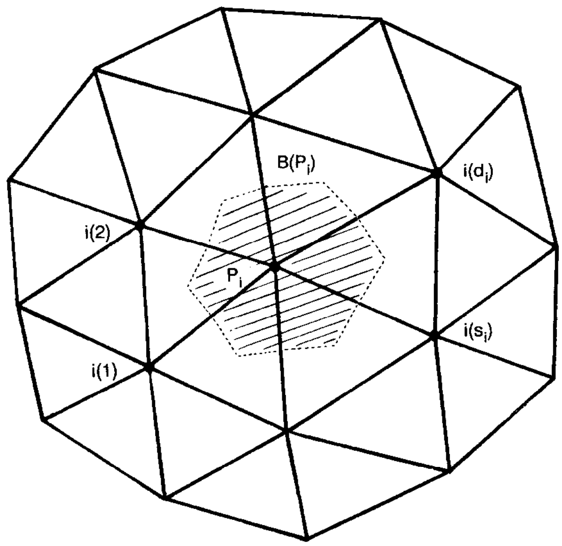

Fig. 1. Triangulation.

Thus, a triangulation $\mathscr{T}^h$ is established (see Fig. 1). We say that a family $\{\mathscr{T}^h\}$ of triangulations is *regular* if there exists a positive constant c independent of the triangulation such that

$$h_q \leqslant c\beta_q, \quad \text{for all } T_q \in \mathscr{T}^h.$$

Remark 2.1. In the case $n = 2$, $\{\mathscr{T}^h\}$ is regular if there exists a constant θ_0 satisfying $0 < \theta_0 \leqslant \theta_{\min}$, where $\theta_{\min}$ denotes the smallest angle of all the triangles $T_q \in \mathscr{T}^h$.

For an n-simplex $T_q \in \mathscr{T}^h$, let $P_0^{(q)} = P_i$, $P_1^{(q)} = P_{i_1}, \ldots, P_n^{(q)} = P_{i_n}$ be its vertices, and let $\lambda_j^{(q)}(x)$, $0 \leqslant j \leqslant n$, be the barycentric coordinates of a point $x \in T_q$ with respect to $P_j^{(q)}$, $0 \leqslant j \leqslant n$, respectively. Define the lumped mass region $\mathscr{B}(P_i)$ corresponding to P_i as follows (see Fig. 1):

$$\mathscr{B}(P_i) = \bigcup_q \left\{ B_i^q;\ T_q \in \mathscr{T}^h \text{ such that } P_i \text{ is a vertex of } T_q \right\}, \quad 1 \leqslant i \leqslant N + M,$$

with

$$B_i^q = \bigcap_{j=1}^{n} \left\{ x \in T_q;\ \lambda_0^{(q)}(x) \geqslant \lambda_j^{(q)}(x) \right\}.$$

Let $\phi_{h,i}$, $\psi_{h,i}$, $1 \leqslant i \leqslant N + M$, be the finite-element basis such that $\phi_{h,i}$ is continuous in $\overline{\Omega}$ and linear on each T_q,

$$\phi_{h,i}(P_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

$$\psi_{h,i}(x) = \begin{cases} 1, & x \in \mathscr{B}(P_i), \\ 0, & x \notin \mathscr{B}(P_i), \end{cases}$$

for $1 \leqslant i,\ j \leqslant N + M$. The use of $\{\psi_{h,i}\}$ associated with the lumped mass region results in a significant reduction in computational efforts. If we seek a finite-element solution $\{u_h, v_h\}$ for (1.1) in the form

$$\begin{aligned} u_h &= \sum_{i=1}^{N} \xi_i \phi_{h,i} + \sum_{i=N+1}^{N+M} g_{1,i} \phi_{h,i}, \\ v_h &= \sum_{i=1}^{N} \zeta_i \phi_{h,i} + \sum_{i=N+1}^{N+M} g_{2,i} \phi_{h,i}, \end{aligned} \tag{2.1}$$

with

$$\begin{aligned} &\xi_i = u_h(P_i), \quad \zeta_i = v_h(P_i), \quad && 1 \leqslant i \leqslant N, \\ &g_{1,i} = g_1(P_i), \quad g_{2,i} = g_2(P_i), \quad && N+1 \leqslant i \leqslant N+M, \end{aligned}$$

then the solution vector $(\boldsymbol{\xi}, \boldsymbol{\zeta}) = (\xi_1, \ldots, \xi_N, \zeta_1, \ldots, \zeta_N)$ satisfies the following system of nonlinear equations:

$$\begin{aligned} H_i(\boldsymbol{\xi}, \boldsymbol{\zeta}) &\equiv \sum_{j=1}^{N} a_{i,j} \xi_j + \sum_{j=N+1}^{N+M} a_{i,j} g_{1,j} + b_1 m_i \xi_i^{n_1} \zeta_i^{n_2} = 0, \quad 1 \leqslant i \leqslant N, \\ Q_i(\boldsymbol{\xi}, \boldsymbol{\zeta}) &\equiv \sum_{j=1}^{N} a_{i,j} \zeta_j + \sum_{j=N+1}^{N+M} a_{i,j} g_{2,j} + b_2 m_i \xi_i^{n_1} \zeta_i^{n_2} = 0, \quad 1 \leqslant i \leqslant N. \end{aligned} \tag{2.2}$$

Here

$$a_{i,j} = \sum_{r=1}^{n} \int_\Omega \frac{\partial \phi_{h,j}}{\partial x_r} \frac{\partial \phi_{h,i}}{\partial x_r} \mathrm{d}x, \quad 1 \leqslant i \leqslant N, \ 1 \leqslant j \leqslant N+M, \tag{2.3}$$

$$m_i = \int_\Omega \psi_{h,i}^2 \, \mathrm{d}x, \quad 1 \leqslant i \leqslant N. \tag{2.4}$$

In the sequel, we use matrix notations [2,16,18,24]. Given a matrix $\boldsymbol{B} = (b_{i,j})$, $1 \leqslant i, \ j \leqslant \overline{N}$, we say that $\boldsymbol{B}$ is *reducible* if there exists a permutation matrix $\boldsymbol{P}$ such that

$$\boldsymbol{P}\boldsymbol{B}\boldsymbol{P}^{\mathrm{t}} = \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \boldsymbol{0} & \boldsymbol{B}_3 \end{pmatrix},$$

where t denotes the transpose, $\boldsymbol{B}_1$ is an $r \times r$ submatrix and $\boldsymbol{B}_3$ is an $(\overline{N} - r) \times (\overline{N} - r)$ submatrix with $1 \leqslant r < \overline{N}$. If no such permutation matrix exists, then $\boldsymbol{B}$ is *irreducible*. We also say that $\boldsymbol{B}$ is *diagonally dominant* if

$$|b_{i,i}| \geqslant \sum_{\substack{j=1 \\ j \neq i}}^{N} |b_{i,j}|, \quad 1 \leqslant i \leqslant \overline{N}. \tag{2.5}$$

Furthermore, $\boldsymbol{B}$ is *irreducible diagonally dominant* if it is irreducible and diagonally dominant and the strict inequality in (2.5) holds for at least one i. Let μ_i, $1 \leqslant i \leqslant \overline{N}$, be the eigenvalues of $\boldsymbol{B}$. The spectra radius of $\boldsymbol{B}$ is defined by

$$\rho(\boldsymbol{B}) = \max_{1 \leqslant i \leqslant \overline{N}} |\mu_i|.$$

Let $\boldsymbol{B} = \boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U}$, where $\boldsymbol{D}$, $-\boldsymbol{L}$ and $-\boldsymbol{U}$ are the diagonal, strictly lower triangular and strictly upper triangular parts of $\boldsymbol{B}$, respectively, and $\boldsymbol{D}$ is invertible. Then $\boldsymbol{B}$ is *consistently* *ordered* if the eigenvalues of $\alpha D^{-1}L + \alpha^{-1}D^{-1}U$ are independent of α for all $\alpha \neq 0$. Moreover, B is 2-*cyclic* if there is a permutation matrix P such that

$$PD^{-1}(L+U)P^{t} = \begin{pmatrix} 0 & B_1 \\ B_2 & 0 \end{pmatrix},$$

where the diagonal blocks are square. For $B=(b_{i,j})$, $C=(c_{i,j})$, $1 \leqslant i,\ j \leqslant \overline{N}$, we write $B \geqslant C$ (or $B \geqslant 0$) if $b_{i,j} \geqslant c_{i,j}$ (or $b_{i,j} \geqslant 0$), $1 \leqslant i,\ j \leqslant \overline{N}$. Similarly, for $y=(y_1, y_2, \ldots, y_{\overline{N}})^{t}$, $z=(z_1, z_2, \ldots, z_{\overline{N}})^{t}$, we write $y \geqslant z$ (or $y \geqslant 0$) if $y_i \geqslant z_i$ (or $y_i \geqslant 0$), $1 \leqslant i \leqslant \overline{N}$. We define $|B| = (|b_{i,j}|)$, $1 \leqslant i,\ j \leqslant \overline{N}$, $|y| = (|y_1|, |y_2|, \ldots, |y_{\overline{N}}|)^{t}$, and the comparison matrix $\mathscr{M}(B)$ by

$$(\mathscr{M}(B))_{i,i} = |b_{i,i}|, \quad (\mathscr{M}(B))_{i,j} = -|b_{i,j}|, \qquad i \neq j, \quad 1 \leqslant i,\ j \leqslant \overline{N}.$$

B is an *M-matrix* if $b_{i,j} \leqslant 0$, $i \neq j$, $1 \leqslant i,\ j \leqslant \overline{N}$, and the inverse B^{-1} exists and $B^{-1} \geqslant 0$. Also B is an *H-matrix* if $\mathscr{M}(B)$ is an M-matrix [2, p.184].

According to [3,4], we say that a triangulation $\mathscr{T}^h$ is of *nonnegative type* if its corresponding matrix $\tilde{A} = (a_{i,j})$, $1 \leqslant i \leqslant N$, $1 \leqslant j \leqslant N+M$, defined by (2.3) has the following properties:

(i) $$a_{i,i} > 0, \quad a_{i,j} \leqslant 0, \qquad i \neq j, \quad 1 \leqslant i \leqslant N,\ 1 \leqslant j \leqslant N+M,$$
$$\sum_{j=1}^{N+M} a_{i,j} \geqslant 0, \quad 1 \leqslant i \leqslant N,$$

(ii) $A = (a_{i,j})$, $1 \leqslant i,\ j \leqslant N$, is irreducibly diagonally dominant.

Remark 2.2. In the case $n=2$, $\mathscr{T}^h$ is of nonnegative type if all the angles of the triangles are less than or equal to $\frac{1}{2}\pi$ [4]. Moreover, the matrix A is consistently ordered and 2-cyclic if the nodal points are numbered left to right, bottom to top, calling this the natural ordering of the nodal points [24, p.187].

Now we make the following assumptions.

Assumption 2.3. The triangulation $\mathscr{T}^h$ is of nonnegative type and regular.

Assumption 2.4. The matrix A is consistently ordered and 2-cyclic.

Remark 2.5. In view of (1.2), there exist $1 \leqslant i_1,\ i_2 \leqslant N$, $N+1 \leqslant j_1,\ j_2 \leqslant N+M$ such that $a_{i_1,j_1} < 0$, $g_{1,j_1} > 0$, $a_{i_2,j_2} < 0$, $g_{2,j_2} > 0$, under Assumption 2.3 with sufficiently small h. Hence we can avoid the discrete trivial solution ($\xi_i = 0$, $1 \leqslant i \leqslant N$ or $\zeta_i = 0$, $1 \leqslant i \leqslant N$). Further, Assumption 2.3 yields the discrete maximum principles [3,4,9]. Thus, under Assumption 2.3 with sufficiently small h, any nonnegative solution $(\xi_1, \ldots, \xi_N, \zeta_1, \ldots, \zeta_N)$ of (2.2) satisfies the following constraint:

$$0 < \xi_i < G_1, \quad 0 < \zeta_i < G_2, \qquad 1 \leqslant i \leqslant N,$$

with

$$G_1 \equiv \max_{N+1 \leqslant j \leqslant N+M} g_{1,j}, \qquad G_2 \equiv \max_{N+1 \leqslant j \leqslant N+M} g_{2,j}.$$

This constraint is the discrete counterpart of (1.3).

Next, we present the projected SOR method. Define intervals I_1 and I_2 by

$$I_1 = [0,\ G_1], \qquad I_2 = [0,\ G_2],$$

and define projections $\mathscr{P}_{(1)} : \mathbb{R} \to I_1$ and $\mathscr{P}_{(2)} : \mathbb{R} \to I_2$ by

$$\mathscr{P}_{(1)}(y) = \min\{G_1,\ \max\{0,\ y\}\},$$

$$\mathscr{P}_{(2)}(y) = \min\{G_2,\ \max\{0,\ y\}\}.$$

It is noted that $\mathscr{P}_{(1)}(y)$ and $\mathscr{P}_{(2)}(y)$ are continuous. Then we have the following simple algorithm for solving the nonlinear equations (2.2):

$$\begin{aligned} \xi_{i,k+1} &= \mathscr{P}_{(1)}\left(\xi_{i,k} - \omega \frac{H_i(\boldsymbol{\xi}_k^{i-1},\ \boldsymbol{\zeta}_k^0)}{\partial_i H_i(\boldsymbol{\xi}_k^{i-1},\ \boldsymbol{\zeta}_k^0)} \right), \quad 1 \leqslant i \leqslant N, \\ \zeta_{i,k+1} &= \mathscr{P}_{(2)}\left(\zeta_{i,k} - \omega \frac{Q_i(\boldsymbol{\xi}_{k+1}^0,\ \boldsymbol{\zeta}_k^{i-1})}{\partial_i Q_i(\boldsymbol{\xi}_{k+1}^0,\ \boldsymbol{\zeta}_k^{i-1})} \right), \quad 1 \leqslant i \leqslant N, \end{aligned} \tag{2.6}$$

for $k = 0, 1, 2, \ldots$. Here ω is a relaxation parameter such that $0 < \omega < 2$ (see Theorems 3.5 and 4.2 below), and $\xi_{i,0} \in I_1$, $\zeta_{i,0} \in I_2$, $1 \leqslant i \leqslant N$, are initial values and

$$\partial_i H_i = \frac{\partial H_i}{\partial \xi_i}, \qquad \partial_i Q_i = \frac{\partial Q_i}{\partial \zeta_i},$$

$$\boldsymbol{\xi}_k^0 = (\xi_{1,k},\ \xi_{2,k}, \ldots,\ \xi_{N,k})^{\mathrm{t}} = \boldsymbol{\xi}_k,$$

$$\boldsymbol{\zeta}_k^0 = (\zeta_{1,k},\ \zeta_{2,k}, \ldots,\ \zeta_{N,k})^{\mathrm{t}} = \boldsymbol{\zeta}_k,$$

$$\vdots \qquad \vdots$$

$$\boldsymbol{\xi}_k^{i-1} = (\xi_{1,k+1}, \ldots,\ \xi_{i-1,k+1},\ \xi_{i,k}, \ldots,\ \xi_{N,k})^{\mathrm{t}},$$

$$\boldsymbol{\zeta}_k^{i-1} = (\zeta_{1,k+1}, \ldots,\ \zeta_{i-1,k+1},\ \zeta_{i,k}, \ldots,\ \zeta_{N,k})^{\mathrm{t}},$$

$$\vdots \qquad \vdots$$

$$\boldsymbol{\xi}_k^N = (\xi_{1,k+1},\ \xi_{2,k+1}, \ldots,\ \xi_{N,k+1})^{\mathrm{t}} = \boldsymbol{\xi}_{k+1}^0 = \boldsymbol{\xi}_{k+1},$$

$$\boldsymbol{\zeta}_k^N = (\zeta_{1,k+1},\ \zeta_{2,k+1}, \ldots,\ \zeta_{N,k+1})^{\mathrm{t}} = \boldsymbol{\zeta}_{k+1}^0 = \boldsymbol{\zeta}_{k+1}.$$

In [8], we presented the explicit monotone iteration

$$\begin{aligned} \bar{w}_{i,k+1} &= \bar{w}_{i,k} - \frac{H_i(\bar{\boldsymbol{w}}_k^{i-1},\ \bar{\boldsymbol{y}}_k^0)}{a_{i,i} + b_1 m_i \bar{y}_{i,k}^{n_2} F[n_1 - 1;\ \bar{w}_{i,k},\ \bar{w}_{i,k}^{\max}]}, \quad 1 \leqslant i \leqslant N, \\ \bar{y}_{i,k+1} &= \bar{y}_{i,k} - \frac{Q_i(\bar{\boldsymbol{w}}_{k+1}^0,\ \bar{\boldsymbol{y}}_k^{i-1})}{a_{i,i} + b_2 m_i \bar{w}_{i,k+1}^{n_1} F[n_2 - 1;\ \bar{y}_{i,k},\ \bar{y}_{i,k}]}, \quad 1 \leqslant i \leqslant N, \end{aligned} \tag{2.7}$$

for $k = 0, 1, 2, \dots$. Here

$$\bar{w}_{i,0} = 0, \quad \bar{y}_{i,0} = G_2, \qquad 1 \leqslant i \leqslant N,$$

$$F[s;\ y,\ z] = \begin{cases} \sum_{j=0}^{s} y^{s-j} z^j, & s \geqslant 1, \\ 1, & s = 0, \end{cases} \tag{2.8}$$

$$\bar{w}_{i,k}^{\max} = \max\{\bar{w}_{i(1),k+1}, \bar{w}_{i(2),k+1}, \dots, \bar{w}_{i(d_i),k+1}, \bar{w}_{i(d_i+1),k}, \dots, \bar{w}_{i(s_i),k}\},$$

$i(1),\ i(2), \dots, i(d_i),\ i(d_i+1), \dots, i(s_i)$ are nodal numbers of the vertices $P_{i(1)}, P_{i(2)}, \dots, P_{i(d_i)}, P_{i(d_i+1)}, \dots, P_{i(s_i)}$ associated with P_i such that $\overline{P_i P_{i(j)}}$, $1 \leqslant j \leqslant s_i$, are sides of some n-simplices of $\mathscr{T}^h$ and

$$i(1) < i,\ i(2) < i, \dots, i(d_i) < i,\ i(d_i+1) > i, \dots, i(s_i) > i,$$

$$a_{i,i(j)} < 0, \quad 1 \leqslant j \leqslant s_i.$$

See Fig. 1. As pointed out, in the iterations (2.6) and (2.7), one need not solve a system of linear equations at each stage. This is a desirable feature from a computational point of view. However, the speed of convergence in (2.7) is slow for a large scale problem.

We end this section by stating some general results on matrices, which are useful later.

Lemma 2.6 (Ortega [16, p.110]). *Let $\boldsymbol{B} = \boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U} = (b_{i,j})$, $1 \leqslant i,\ j \leqslant \bar{N}$, be irreducibly diagonally dominant with $b_{i,i} > 0$, $b_{i,j} \leqslant 0$, $i \neq j$, $1 \leqslant i,\ j \leqslant \bar{N}$, where $\boldsymbol{D}$, $-\boldsymbol{L}$ and $-\boldsymbol{U}$ are the diagonal, strictly lower triangular and strictly upper triangular parts of $\boldsymbol{B}$, respectively. Then $\boldsymbol{B}$ is an M-matrix and $\rho(\boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U})) < 1$.*

Lemma 2.7 (Ortega and Rheinboldt [18, p.53]). *Let $\boldsymbol{D}$ and $\boldsymbol{L}$ be $\bar{N} \times \bar{N}$ matrices. If $\boldsymbol{D} \geqslant \boldsymbol{0}$, $\boldsymbol{L} \geqslant \boldsymbol{0}$, $\boldsymbol{D}$ is diagonal and invertible and $\boldsymbol{L}$ is strictly lower triangular, then $(\boldsymbol{D} - \boldsymbol{L})^{-1} \geqslant \boldsymbol{0}$.*

Lemma 2.8 (Pang [21]). *Let $\boldsymbol{B} = \boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U}$ be an $\bar{N} \times \bar{N}$ matrix with all positive diagonal elements, where $\boldsymbol{D}$, $-\boldsymbol{L}$ and $-\boldsymbol{U}$ are the diagonal, strictly lower triangular and strictly upper triangular parts of $\boldsymbol{B}$, respectively. Then, $\boldsymbol{B}$ is an H-matrix if and only if*

$$\rho\left(\left(\frac{1}{\omega}\boldsymbol{D} - |\boldsymbol{L}|\right)^{-1}\left(|1 - \frac{1}{\omega}|\boldsymbol{D} + |\boldsymbol{U}|\right)\right) < 1,$$

for all $0 < \omega < 2/(1 + \rho(|\boldsymbol{J}|))$, where $\boldsymbol{J} = \boldsymbol{D}^{-1}(|\boldsymbol{L}| - |\boldsymbol{U}|)$.

Lemma 2.9 (Ortega [16, p.25]). *Let $\boldsymbol{B}$ be an $\bar{N} \times \bar{N}$ matrix. Then $\lim_{k \to \infty} \boldsymbol{B}^k = \boldsymbol{0}$ if and only if $\rho(\boldsymbol{B}) < 1$.*

Lemma 2.10 (Ortega [16, p.45]). *The eigenvalues of a matrix are continuous functions of the elements of the matrix.*

Lemma 2.11 (Ortega [16], Varga [24], Young [25]). *Let $\boldsymbol{B} = \boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U}$ be a symmetric M-matrix, where $\boldsymbol{D}$, $-\boldsymbol{L}$ and $-\boldsymbol{U}$ are the diagonal, strictly lower triangular and strictly upper triangular parts* *of $\boldsymbol{B}$, respectively. Assume that $\boldsymbol{B}$ is consistently ordered and 2-cyclic. Then, for the relaxation parameter ω, the SOR iteration matrix*

$$\boldsymbol{H}_\omega \equiv (\boldsymbol{D}-\omega\boldsymbol{L})^{-1}((1-\omega)\boldsymbol{D}+\omega\boldsymbol{U})$$

satisfies

$$\rho(\boldsymbol{H}_\omega)<1, \quad \textit{for all } 0<\omega<2,$$

and there exists the optimum relaxation parameter $\omega_{\text{opt}}^{(\text{L})}$ such that

$$\omega_{\text{opt}}^{(\text{L})}=\frac{2}{\left(1+\sqrt{1-\mu^2}\right)}, \qquad \mu=\rho\left(\boldsymbol{D}^{-1}(\boldsymbol{L}+\boldsymbol{U})\right), \qquad \rho\left(\boldsymbol{H}_{\omega_{\text{opt}}^{(\text{L})}}\right)=\min_{0<\omega<2}\rho(\boldsymbol{H}_\omega).$$

Remark 2.12. As well known [16,24,25], we can show the graph of $\rho(\boldsymbol{H}_\omega)$ as a function of ω in Fig. 2. For details on the determination of $\omega_{\text{opt}}^{(\text{L})}$, see [25, Chapter 6].

3. Convergence result

In this section, we prove the convergence of (2.6). The Jacobian matrix of $(H_1,\ldots,H_N, Q_1,\ldots,Q_N)$ is not symmetric, so that there is no longer an associated minimization problem. Unlike [10,11], our present method of the convergence proof is based upon a contraction argument with matrix theory. For the matrix $\boldsymbol{A}=(a_{i,j})$, $1\leqslant i,\ j\leqslant N$, defined by (2.3), let $\boldsymbol{A}=\boldsymbol{D}-\boldsymbol{L}-\boldsymbol{U}$, where $\boldsymbol{D}$, $-\boldsymbol{L}$ and $-\boldsymbol{U}$ are the diagonal, strictly lower triangular and strictly upper triangular parts of $\boldsymbol{A}$, respectively. Further, set

$$\boldsymbol{D}_k^{(1)}=b_1n_1\,\operatorname{diag}\left(m_1\xi_{1,k}^{n_1-1}\zeta_{1,k}^{n_2},\ m_2\xi_{2,k}^{n_1-1}\zeta_{2,k}^{n_2},\ldots,m_N\xi_{N,k}^{n_1-1}\zeta_{N,k}^{n_2}\right),$$
$$\boldsymbol{D}_k^{(2)}=b_2n_2\,\operatorname{diag}\left(m_1\xi_{1,k+1}^{n_1}\zeta_{1,k}^{n_2-1},\ m_2\xi_{2,k+1}^{n_1}\zeta_{2,k}^{n_2-1},\ldots,m_N\xi_{N,k+1}^{n_1}\zeta_{N,k}^{n_2-1}\right),$$
$$\boldsymbol{A}_k^{(1)}=\frac{1}{\omega}\left(\boldsymbol{D}+\boldsymbol{D}_k^{(1)}\right)-|\boldsymbol{L}|, \qquad \boldsymbol{A}_k^{(2)}=\frac{1}{\omega}\left(\boldsymbol{D}+\boldsymbol{D}_k^{(2)}\right)-|\boldsymbol{L}|,$$
$$\boldsymbol{B}=\frac{1}{\omega}\boldsymbol{D}-|\boldsymbol{L}|, \qquad \boldsymbol{C}=\left|1-\frac{1}{\omega}\right|\boldsymbol{D}+|\boldsymbol{U}|, \qquad \boldsymbol{J}=\boldsymbol{D}^{-1}(|\boldsymbol{L}|-|\boldsymbol{U}|),$$
$$\tilde{\boldsymbol{B}}_\epsilon=\begin{pmatrix}\boldsymbol{B} & \boldsymbol{0}\\ -\epsilon\boldsymbol{I} & \boldsymbol{B}\end{pmatrix}, \qquad \tilde{\boldsymbol{C}}_\epsilon=\begin{pmatrix}\boldsymbol{C}+\epsilon\boldsymbol{I} & \epsilon\boldsymbol{I}\\ \boldsymbol{0} & \boldsymbol{C}+\epsilon\boldsymbol{I}\end{pmatrix},$$

where $\omega>0$, $\epsilon>0$ and $\boldsymbol{I}$ is the $N\times N$ identity matrix.

Combining the result in [7] and Remark 2.5, we have the following theorem.

Theorem 3.1 (Ishihara [7]). *Under Assumption* 2.3 *with sufficiently small h, there exists a unique nonnegative solution $(\bar{\xi},\bar{\zeta})=(\bar{\xi}_1,\ldots,\bar{\xi}_N,\bar{\zeta}_1,\ldots,\bar{\zeta}_N)$ of* (2.2), *which satisfies*

$$0<\bar{\xi}_i<G_1, \quad 0<\bar{\zeta}_i<G_2, \qquad 1\leqslant i\leqslant N.$$

Furthermore, some lemmas are prepared.

Lemma 3.2. *For F defined by* (2.8), *it holds that*

$$y^{n_1}z^{n_2}-a^{n_1}b^{n_2}=z^{n_2}F[n_1-1;\ y,\ a](y-a)+a^{n_1}F[n_2-1;\ z,\ b](z-b).$$

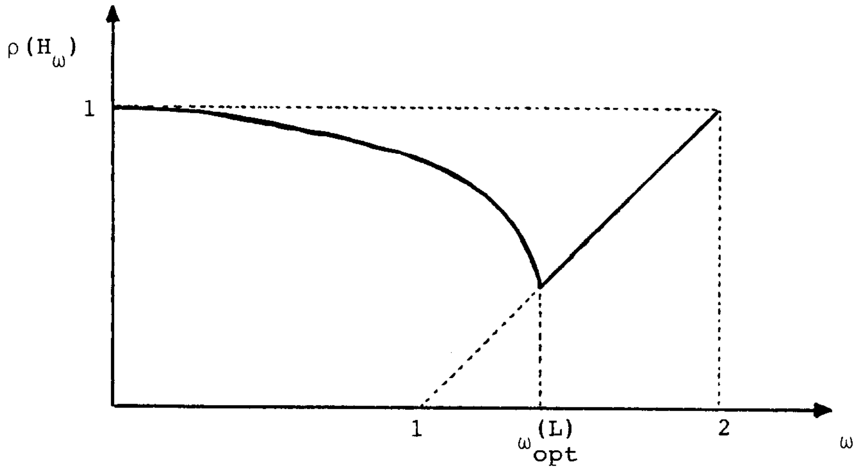

Fig. 2. $\rho(H_\omega)$ for the linear problem with the matrix A.

Proof. A simple computation leads to

$$\begin{aligned} y^{n_1}z^{n_2} - a^{n_1}b^{n_2} &= z^{n_2}(y^{n_1} - a^{n_1}) + a^{n_1}(z^{n_2} - b^{n_2}) \\ &= z^{n_2}F[n_1 - 1;\ y,\ a](y-a) + a^{n_1}F[n_2 - 1;\ z,\ b](z-b). \qquad \square \end{aligned}$$

Lemma 3.3. *Under Assumption* 2.3, *there exists* $1 < \omega^* \leqslant 2$ *such that for all* $0 < \omega < \omega^*$,

$$\rho(B^{-1}C) < 1.$$

Proof. From Assumption 2.3, Lemma 2.6 and the fact that $\mathcal{M}(A) = A$, A is also an H-matrix and $\rho(\,|J|\,) < 1$. An application of Lemma 2.8 gives $\rho(B^{-1}C) < 1$ for all $0 < \omega < 2/(1 + \rho(\,|J|\,))$. By putting $\omega^* = 2/(1 + \rho(\,|J|\,))$, we have $1 < \omega^* \leqslant 2$. This completes the proof. $\square$

Lemma 3.4. *Let* $\omega > 0$. *Under Assumption* 2.3 *with sufficiently small* h, (2.6) *satisfies*:
if $\xi_{i,k+1} \leqslant \bar{\xi}_i$ (*or* $\xi_{i,k+1} \geqslant \bar{\xi}_i$), *then*

$$H_i(\boldsymbol{\xi}_k,\ \boldsymbol{\zeta}_k) + \left(A_k^{(1)}(\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_k)\right)_i \geqslant 0 \quad \left(\text{or } H_i(\boldsymbol{\xi}_k,\ \boldsymbol{\zeta}_k) + \left(A_k^{(1)}(\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_k)\right)_i \leqslant 0\right);$$

if $\zeta_{i,k+1} \leqslant \bar{\zeta}_i$ (*or* $\zeta_{i,k+1} \geqslant \bar{\zeta}_i$), *then*

$$Q_i(\boldsymbol{\xi}_{k+1},\ \boldsymbol{\zeta}_k) + \left(A_k^{(2)}(\boldsymbol{\zeta}_{k+1} - \boldsymbol{\zeta}_k)\right)_i \geqslant 0 \quad \left(\text{or } Q_i(\boldsymbol{\xi}_{k+1},\ \boldsymbol{\zeta}_k) + \left(A_k^{(2)}(\boldsymbol{\zeta}_{k+1} - \boldsymbol{\zeta}_k)\right)_i \leqslant 0\right).$$

Proof. Put

$$V_{i,k} = \frac{H_i(\boldsymbol{\xi}_k^{i-1},\ \boldsymbol{\zeta}_k^0)}{\partial_i H_i(\boldsymbol{\xi}_k^{i-1},\ \boldsymbol{\zeta}_k^0)}.$$

Consider the case $\xi_{i,k+1} \leqslant \bar{\xi}_i$. Assume that $\xi_{i,k} - \omega V_{i,k} > G_1$. Then, from (2.6) and Theorem 3.1 we have $\xi_{i,k+1} = G_1 > \bar{\xi}_i$. This is a contradiction. Consequently it holds that $\xi_{i,k} - \omega V_{i,k} \leqslant G_1$, from which we have

$$\xi_{i,k+1} = \begin{cases} \xi_{i,k} - \omega V_{i,k}, & \text{if } 0 \leqslant \xi_{i,k} - \omega V_{i,k} \leqslant G_1, \\ 0 \geqslant \xi_{i,k} - \omega V_{i,k}, & \text{if } \xi_{i,k} - \omega V_{i,k} \leqslant 0. \end{cases}$$

Hence, we get $\xi_{i,k+1} \geqslant \xi_{i,k} - \omega V_{i,k}$, i.e.,

$$\frac{1}{\omega}\left(a_{i,i} + n_1 b_1 m_i \xi_{i,k}^{n_1-1} \zeta_{i,k}^{n_2}\right)\left(\xi_{i,k+1} - \xi_{i,k}\right) + \sum_{j=1}^{i-1} a_{i,j}\xi_{j,k+1} + \sum_{j=i}^{N} a_{i,j}\xi_{j,k} + \sum_{j=N+1}^{N+M} a_{i,j} g_{1,j} + b_1 m_i \xi_{i,k}^{n_1} \zeta_{i,k}^{n_2} \geqslant 0.$$

This is written as

$$H_i(\xi_k, \zeta_k) + \left(A_k^{(1)}(\xi_{k+1} - \xi_k)\right)_i \geqslant 0.$$

Next, consider the case $\xi_{i,k+1} \geqslant \bar{\xi}_i$. Assume that $\xi_{i,k} - \omega V_{i,k} < \bar{\xi}_i$. Then we have

$$\xi_{i,k+1} = \begin{cases} \xi_{i,k} - \omega V_{i,k} < \bar{\xi}_i, & \text{if } 0 \leqslant \xi_{i,k} - \omega V_{i,k} < \bar{\xi}_i, \\ 0 < \bar{\xi}_i, & \text{if } \xi_{i,k} - \omega V_{i,k} \leqslant 0. \end{cases}$$

This is a contradiction. Consequently it holds that $\xi_{i,k} - \omega V_{i,k} \geqslant \bar{\xi}_i$, from which we have

$$\xi_{i,k+1} = \begin{cases} \xi_{i,k} - \omega V_{i,k}, & \text{if } \bar{\xi}_i \leqslant \xi_{i,k} - \omega V_{i,k} \leqslant G_1, \\ G_1 \leqslant \xi_{i,k} - \omega V_{i,k}, & \text{if } \xi_{i,k} - \omega V_{i,k} \geqslant G_1. \end{cases}$$

Hence, we get $\xi_{i,k+1} \leqslant \xi_{i,k} - \omega V_{i,k}$. This is written as

$$H_i(\xi_k, \zeta_k) + \left(A_k^{(1)}(\xi_{k+1} - \xi_k)\right)_i \leqslant 0.$$

Similarly, we obtain the other results. Thus, the proof is complete. □

Now, we prove our main result.

Theorem 3.5. *Under Assumption* 2.3 *with sufficiently small h, there exists* $1 < \omega^* \leqslant 2$ *such that for all* $0 < \omega < \omega^*$, *the iteration* (2.6) *satisfies*

$$\lim_{k\to\infty} \xi_{i,k} = \bar{\xi}_i, \quad \lim_{k\to\infty} \zeta_{i,k} = \bar{\zeta}_i, \qquad 1 \leqslant i \leqslant N.$$

Proof. Let $\epsilon > 0$ be an arbitrary number. Let Assumption 2.3 with sufficiently small h hold. From (2.3) and (2.4), it follows that

$$\frac{m_i}{a_{i,i}} = \mathcal{O}(h^2), \quad m_i = \mathcal{O}(h^n), \qquad 1 \leqslant i \leqslant N.$$

As a consequence, we can choose m_i small enough in comparison with $a_{i,i}$ and

$$m_i\left(1 + \frac{1}{\omega}\right) \max\{b_1, b_2\} \max\left\{n_1 G_1^{n_1-1} G_2^{n_2},\ n_2 G_1^{n_1} G_2^{n_2-1}\right\} < \epsilon, \quad 1 \leqslant i \leqslant N. \tag{3.1}$$

We evaluate $|\bar{\xi} - \xi_{k+1}|$. Consider any index i such that $\xi_{i,k+1} \leqslant \bar{\xi}_i$. Since $H_i(\bar{\xi}, \bar{\zeta}) = 0$, an application of Lemma 3.4 gives

$$H_i(\xi_k, \zeta_k) + \left(A_k^{(1)}(\xi_{k+1} - \xi_k)\right)_i - H_i(\bar{\xi}, \bar{\zeta}) \geqslant 0.$$

Thus using Lemma 3.2, (3.1), Theorem 3.1 and the fact that $0 \leqslant \xi_{i,k} \leqslant G_1$, $0 \leqslant \zeta_{i,k} \leqslant G_2$, we have

$$\begin{aligned}
\left(\boldsymbol{B}|\bar{\xi}-\xi_{k+1}|\right)_i &\leqslant \left(A_k^{(1)}|\bar{\xi}-\xi_{k+1}|\right)_i \\
&= \left(A_k^{(1)}\right)_{i,i}\left(\bar{\xi}_i-\xi_{i,k+1}\right) - \sum_{j=1}^{i-1}\left|\left(A_k^{(1)}\right)_{i,j}\left(\bar{\xi}_j-\xi_{j,k+1}\right)\right| \\
&\leqslant \left(A_k^{(1)}(\bar{\xi}-\xi_{k+1})\right)_i \leqslant -\left(H_i(\bar{\xi},\bar{\zeta}) - H_i(\xi_k,\zeta_k) - \left(A_k^{(1)}(\bar{\xi}-\xi_k)\right)_i\right) \\
&= -\left(\sum_{j=1}^{N} a_{i,j}\left(\bar{\xi}_j-\xi_{j,k}\right) + b_1 m_i\left(\bar{\xi}_i^{n_1}\bar{\zeta}_i^{n_2} - \xi_{i,k}^{n_1}\zeta_{i,k}^{n_2}\right) - \left(A_k^{(1)}(\bar{\xi}-\xi_k)\right)_i\right) \\
&= -\left(\sum_{j=i}^{N} a_{i,j}\left(\bar{\xi}_j-\xi_{j,k}\right) + b_1 m_i \bar{\zeta}_i^{n_2} F\left[n_1-1;\ \bar{\xi}_i,\ \xi_{i,k}\right]\left(\bar{\xi}_i-\xi_{i,k}\right)\right. \\
&\qquad + b_1 m_i \xi_{i,k}^{n_1} F\left[n_2-1;\ \bar{\zeta}_i,\ \zeta_{i,k}\right]\left(\bar{\zeta}_i-\zeta_{i,k}\right) \\
&\qquad \left. - \frac{1}{\omega}\left(a_{i,i} + n_1 b_1 m_i \xi_{i,k}^{n_1-1}\zeta_{i,k}^{n_2}\right)\left(\bar{\xi}_i-\xi_{i,k}\right)\right) \\
&\leqslant \left(\left|1-\frac{1}{\omega}\right| a_{i,i} + \epsilon\right)\left|\bar{\xi}_i-\xi_{i,k}\right| + \sum_{j=i+1}^{N} |a_{i,j}|\left|\bar{\xi}_j-\xi_{j,k}\right| + \epsilon\left|\bar{\zeta}_i-\zeta_{i,k}\right|,
\end{aligned} \tag{3.2}$$

for sufficiently small h.

Next, consider any index i such that $\xi_{i,k+1} \geqslant \bar{\xi}_i$. An application of Lemma 3.4 gives

$$H_i(\xi_k,\zeta_k) + \left(A_k^{(1)}(\xi_{k+1}-\xi_k)\right)_i - H_i(\bar{\xi},\bar{\zeta}) \leqslant 0.$$

Then we have

$$\begin{aligned}
\left(\boldsymbol{B}|\bar{\xi}-\xi_{k+1}|\right)_i &\leqslant \left(A_k^{(1)}|\bar{\xi}-\xi_{k+1}|\right)_i \\
&= -\left(A_k^{(1)}\right)_{i,i}\left(\bar{\xi}_i-\xi_{i,k+1}\right) - \sum_{j=1}^{i-1}\left|\left(A_k^{(1)}\right)_{i,j}\left(\bar{\xi}_j-\xi_{j,k+1}\right)\right| \\
&\leqslant -\left(A_k^{(1)}(\bar{\xi}-\xi_{k+1})\right)_i \leqslant H_i(\bar{\xi},\bar{\zeta}) - H_i(\xi_k,\zeta_k) - \left(A_k^{(1)}(\bar{\xi}-\xi_k)\right)_i,
\end{aligned}$$

so that (3.2) holds. Consequently, we obtain

$$\boldsymbol{B}|\bar{\xi}-\xi_{k+1}| \leqslant (\boldsymbol{C}+\epsilon\boldsymbol{I})|\bar{\xi}-\xi_k| + \epsilon|\bar{\zeta}-\zeta_k|. \tag{3.3}$$

Similarly, keeping in mind the term $Q_i(\xi_{k+1},\zeta_k)$ and (3.1), we get

$$\begin{aligned}
\left(\boldsymbol{B}|\bar{\zeta}-\zeta_{k+1}|\right)_i &\leqslant \left(A_k^{(2)}|\bar{\zeta}-\zeta_k|\right)_i \\
&\leqslant \left(\left|1-\frac{1}{\omega}\right| a_{i,i} + \epsilon\right)\left|\bar{\zeta}_i-\zeta_{i,k}\right| + \sum_{j=i+1}^{N} |a_{i,j}|\left|\bar{\zeta}_j-\zeta_{j,k}\right| \\
&\qquad + \epsilon\left|\bar{\xi}_i-\xi_{i,k+1}\right|,
\end{aligned}$$

from which follows

$$-\epsilon|\bar{\xi}-\xi_{k+1}|+B|\bar{\zeta}-\zeta_{k+1}|\leqslant(C+\epsilon I)|\bar{\zeta}-\zeta_k|. \tag{3.4}$$

Therefore, (3.3) and (3.4) are written as

$$\tilde{B}_\epsilon\begin{pmatrix}|\bar{\xi}-\xi_{k+1}|\\|\bar{\zeta}-\zeta_{k+1}|\end{pmatrix}\leqslant\tilde{C}_\epsilon\begin{pmatrix}|\bar{\xi}-\xi_k|\\|\bar{\zeta}-\zeta_k|\end{pmatrix}.$$

By using Lemma 2.7, we have $\tilde{B}_\epsilon^{-1}\geqslant\mathbf{0}$. Hence it follows that

$$\begin{pmatrix}|\bar{\xi}-\xi_{k+1}|\\|\bar{\zeta}-\zeta_{k+1}|\end{pmatrix}\leqslant\tilde{B}_\epsilon^{-1}\tilde{C}_\epsilon\begin{pmatrix}|\bar{\xi}-\xi_k|\\|\bar{\zeta}-\zeta_k|\end{pmatrix}\leqslant\cdots\leqslant\left(\tilde{B}_\epsilon^{-1}\tilde{C}_\epsilon\right)^{k+1}\begin{pmatrix}|\bar{\xi}-\xi_0|\\|\bar{\zeta}-\zeta_0|\end{pmatrix}. \tag{3.5}$$

From Lemma 2.10, we have

$$\rho\left(\tilde{B}_\epsilon^{-1}\tilde{C}_\epsilon\right)\to\rho\left(B^{-1}C\right)\quad\text{as } \epsilon\to 0.$$

By using Lemma 3.3 under Assumption 2.3 with sufficiently small h, we get $\rho(\tilde{B}_\epsilon^{-1}\tilde{C}_\epsilon)<1$. Hence, from Lemma 2.9 and (3.5), there exists $1<\omega^*\leqslant 2$ such that

$$\lim_{k\to\infty}\xi_{i,k}=\bar{\xi}_i,\quad\lim_{k\to\infty}\zeta_{i,k}=\bar{\zeta}_i,\qquad 1\leqslant i\leqslant N,$$

for all $0<\omega<\omega^*$. This completes the proof. □

Remark 3.6. It holds that

$$\tilde{B}_\epsilon^{-1}\tilde{C}_\epsilon\geqslant\begin{pmatrix}B^{-1}C & \mathbf{0}\\ \mathbf{0} & B^{-1}C\end{pmatrix}\geqslant\mathbf{0},\qquad\rho\left(\tilde{B}_\epsilon^{-1}\tilde{C}_\epsilon\right)\geqslant\rho\left(B^{-1}C\right).$$

Remark 3.7. Let $\{u, v\}$ be the solution of (1.1), and let $\{u_h, v_h\}$ be the finite-element solution defined by (2.1). In [7], we showed that

$$\|u-u_h\|_{L^\infty(\Omega)}=\mathcal{O}(h),\qquad\|v-v_h\|_{L^\infty(\Omega)}=\mathcal{O}(h),$$

i.e.,

$$|u(P_i)-\bar{\xi}_i|=\mathcal{O}(h),\quad|v(P_i)-\bar{\zeta}_i|=\mathcal{O}(h),\qquad 1\leqslant i\leqslant N,$$

under appropriate assumptions.

4. Optimum relaxation parameter

In this section, we consider the rate of convergence for the iteration of (2.6). Following [18], we define the convergence factor $R_1(\bar{\xi},\bar{\zeta})$ of (2.6) at the solution $(\bar{\xi},\bar{\zeta})$ by

$$R_1(\bar{\xi},\bar{\zeta})\equiv\sup\left\{\limsup_{k\to\infty}\|(\xi_k,\zeta_k)-(\bar{\xi},\bar{\zeta})\|;\ (\xi_k,\zeta_k)\in\mathscr{C}\right\},$$

where $\mathscr{C}$ is the set of all sequences $\{(\xi_k,\zeta_k)\}$ generated by (2.6) which converge to $(\bar{\xi},\bar{\zeta})$, and $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^N\times\mathbb{R}^N$. It is well known [18, Chapter 9] that the smaller $R_1(\bar{\xi},\bar{\zeta})$

is, the faster the rate of convergence is. Let $\tilde{D} - \tilde{L}$ and $-\tilde{U}$ be the diagonal, strictly lower triangular and strictly upper triangular parts of the Jacobian matrix of (2.2) at $(\bar{\xi}, \bar{\zeta})$. For $\omega > 0$, let

$$\tilde{H}_\omega(\bar{\xi}, \bar{\zeta}) = (\tilde{D} - \omega\tilde{L})^{-1}((1-\omega)\tilde{D} + \omega\tilde{U}).$$

We prepare the following lemma.

Lemma 4.1. *Assume that* $\rho(\tilde{H}_\omega(\bar{\xi}, \bar{\zeta})) < 1$. *Then, under Assumption* 2.3 *with sufficiently small* h, *we have* $R_1(\bar{\xi}, \bar{\zeta}) = \rho(\tilde{H}_\omega(\bar{\xi}, \bar{\zeta}))$.

Proof. From Theorems 3.1 and 3.5, it follows that for any open neighborhood $S \subset I_1^N \times I_2^N$ of $(\bar{\xi}, \bar{\zeta})$, there exists a natural number k_0 such that $(\xi_k, \zeta_k) \in S$ for all $k \geqslant k_0$. Thus, for all $k \geqslant k_0$, (2.6) is written as

$$\xi_{i,k+1} = \xi_{i,k} - \frac{H_i(\xi_k^{i-1}, \zeta_k^0)}{\partial_i H_i(\xi_k^{i-1}, \zeta_k^0)}, \qquad 1 \leqslant i \leqslant N,$$

$$\zeta_{i,k+1} = \zeta_{i,k} - \frac{Q_i(\xi_{k+1}^0, \zeta_k^{i-1})}{\partial_i Q_i(\xi_{k+1}^0, \zeta_k^{i-1})}, \qquad 1 \leqslant i \leqslant N.$$

Hence, as an immediate consequence of [18, p.323], we obtain $R_1(\bar{\xi}, \bar{\zeta}) = \rho(\tilde{H}_\omega(\bar{\xi}, \bar{\zeta}))$. This completes the proof. □

We are now in a position to conclude the existence of the optimum relaxation parameter.

Theorem 4.2. *Let* $0 < \omega < 2$. *Under Assumptions* 2.3 *and* 2.4 *with sufficiently small* h, *there exists the optimum relaxation parameter* $\omega_{\text{opt}}^{(\text{N})}$ *which minimizes the convergence factor* $R_1(\bar{\xi}, \bar{\zeta})$ *of* (2.6) *such that* $\omega_{\text{opt}}^{(\text{N})} \doteqdot \omega_{\text{opt}}^{(\text{L})}$, *where* $\omega_{\text{opt}}^{(\text{L})}$ *is the optimum relaxation parameter for the linear problem with the matrix* $A = (a_{i,j})$, $1 \leqslant i,\ j \leqslant N$, *defined by* (2.3).

Proof. It is noted that under Assumption 2.3, the matrix A is a symmetric M-matrix. From Lemmas 2.6, 2.10, 2.11, 4.1 and Assumption 2.4, we have

$$R_1(\bar{\xi}, \bar{\zeta}) = \rho(\tilde{H}_\omega(\bar{\xi}, \bar{\zeta})) \doteqdot \rho(H_\omega) < 1, \quad \text{for sufficiently small } h,$$

where H_ω is the SOR iteration matrix for the linear problem with the matrix A, and there exists $\omega_{\text{opt}}^{(\text{N})}$ which minimizes $R_1(\bar{\xi}, \bar{\zeta})$ such that $\omega_{\text{opt}}^{(\text{N})} \doteqdot \omega_{\text{opt}}^{(\text{L})}$. The proof is complete. □

5. Numerical result

In this section, we show some numerical results in the two-dimensional case. Let Ω_{H} and Ω_{S} be a regular hexagon and a square given by

$$\Omega_{\text{H}} = \{(x_1, x_2) \in \mathbb{R}^2;\ \tfrac{1}{4}\sqrt{3} < \sqrt{3}x_1 + x_2 < \tfrac{5}{4}\sqrt{3},\ 0 < x_2 < \tfrac{1}{2}\sqrt{3},$$
$$-\tfrac{3}{4}\sqrt{3} < x_2 - \sqrt{3}x_1 < \tfrac{1}{4}\sqrt{3}\},$$

$$\Omega_{\text{S}} = \{(x_1, x_2) \in \mathbb{R}^2;\ 0 < x_1 < 1,\ 0 < x_2 < 1\}.$$

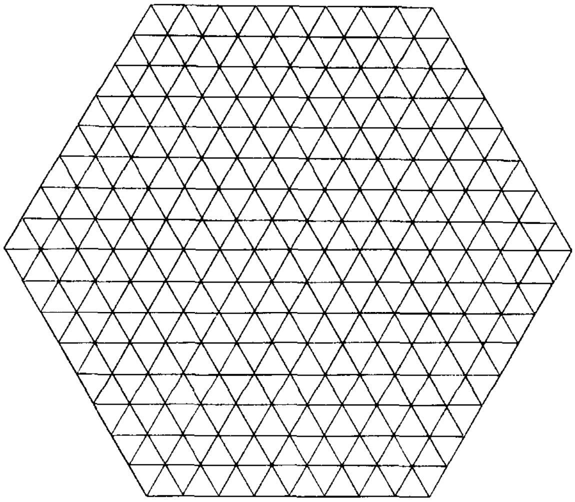

Fig. 3. Uniform mesh of hexagon (217 nodes).

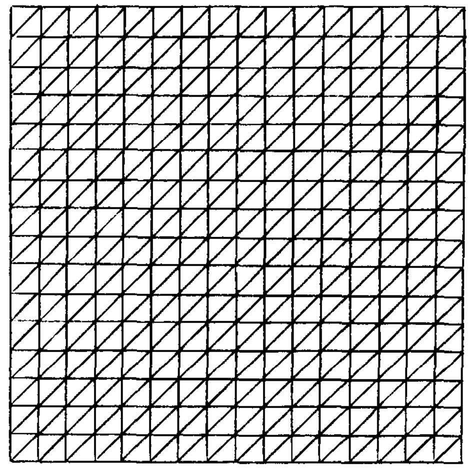

Fig. 4. Uniform mesh of square (289 nodes).

By Γ_{H} and Γ_{S}, we denote the boundaries of Ω_{H} and Ω_{S}, respectively. We give the following problems.

Problem 5.1.

$$\Delta u = 3uv, \qquad \Delta v = 12uv, \qquad \text{in } \Omega_{\mathrm{H}},$$
$$u = \frac{1}{(x_1 + x_2 + 1.75)^2}, \quad v = \frac{4}{(x_1 + x_2 + 1.75)^2}, \qquad \text{on } \Gamma_{\mathrm{H}}.$$

Problem 5.2.

$$\Delta u = 2u^2 v, \qquad \Delta v = u^2 v, \qquad \text{in } \Omega_{\mathrm{S}},$$
$$u = \frac{2}{x_1 + x_2 + 2}, \quad v = \frac{1}{x_1 + x_2 + 2}, \qquad \text{on } \Gamma_{\mathrm{S}}.$$

The exact solutions for Problems 5.1 and 5.2 are the boundary functions extended to the domains $\bar{\Omega}_{\mathrm{H}}$ and $\bar{\Omega}_{\mathrm{S}}$, respectively.

We divide Ω_{H} into uniform mesh with equilateral triangles as shown in Fig. 3 (217 nodes, $h = \frac{1}{16}$, $m_i = \frac{1}{512}\sqrt{3}$, $a_{i,i} = 2\sqrt{3}$, $1 \leqslant i \leqslant N$). We also divide Ω_{S} into a uniform mesh with right isosceles triangles as shown in Fig. 4 (289 nodes, $h = \frac{1}{16}\sqrt{2}$, $m_i = \frac{1}{256}$, $a_{i,i} = 4$, $1 \leqslant i \leqslant N$). The nodal points are numbered with the natural ordering. Thus, from Remark 2.2, Assumptions 2.3 and 2.4 are satisfied, and Remark 2.5 is valid, so that Theorems 3.1, 3.5 and 4.2 hold. For initial data of (2.6), we employ $\xi_{i,0} = 0$, $\zeta_{i,0} = G_2$, $1 \leqslant i \leqslant N$, in comparison with (2.7). We also deal with the system of linear equations

$$\boldsymbol{A}\boldsymbol{y} = \boldsymbol{d}, \quad \text{with } d_i = \sum_{j=1}^{N} a_{i,j}, \ 1 \leqslant i \leqslant N, \tag{5.1}$$

where $\boldsymbol{A} = (a_{i,j})$, $1 \leqslant i$, $j \leqslant N$, is the matrix defined by (2.3), so that its solution is $\boldsymbol{y} =$

Table 1
Number of iterations for Problem 5.1

ω	Number of iterations	
	(2.6)	Linear SOR for (5.1)
1.0	286	275
1.2	194	185
1.4	124	118
1.6	63	58
1.61	60	54
1.62	56	50
1.63	52	50
1.64	50	52
1.65	51	55
1.7	63	67
1.8	96	99
1.9	204	211
(2.7)	286	$\omega_{\text{opt}}^{(\text{L})} \doteqdot 1.62$

Table 2
Number of iterations for Problem 5.2

ω	Number of iterations	
	(2.6)	Linear SOR for (5.1)
1.0	403	423
1.2	273	287
1.4	176	185
1.6	95	100
1.65	73	77
1.66	68	71
1.67	61	64
1.68	63	64
1.69	64	65
1.7	65	65
1.8	97	97
1.9	178	194
(2.7)	403	$\omega_{\text{opt}}^{(\text{L})} \doteqdot 1.67$

$(1, 1, \ldots, 1)^{\text{t}}$. Equation (5.1) is solved by the linear SOR iteration with initial data $(0, 0, \ldots, 0)^{\text{t}}$. The linear optimum relaxation parameters associated with Problems 5.1 and 5.2 are $\omega_{\text{opt}}^{(\text{L})} \doteqdot 1.62$ and $\omega_{\text{opt}}^{(\text{L})} \doteqdot 1.67$, respectively. The numerical convergence criterion is

$$\max\Big\{ \max_{1 \leqslant i \leqslant N} |\xi_{i,k+1} - \xi_{i,k}|, \max_{1 \leqslant i \leqslant N} |\zeta_{i,k+1} - \zeta_{i,k}| \Big\} \leqslant \tfrac{1}{2} \times 10^{-8}.$$

In Tables 1 and 2, we present the comparative numbers of iterations necessary to obtain the numerical convergence criterion in order to determine the optimum relaxation parameter $\omega_{\text{opt}}^{(\text{N})}$. Tables 3 and 4 give finite-element solutions of Problems 5.1 and 5.2 at the nodal points $(0.5, \frac{1}{4}\sqrt{3})$ and $(0.5, 0.5)$, respectively. We can see that the projected SOR method is more effective for increasing the speed of convergence by choosing the optimum relaxation parameter $\omega_{\text{opt}}^{(\text{N})}$ such that $\omega_{\text{opt}}^{(\text{N})} \doteqdot \omega_{\text{opt}}^{(\text{L})}$.

All computations were performed on the FACOM M-382 computer at Kyushu University by using double-precision arithmetic in FORTRAN which carries about 16 significant digits. All data in Tables 3 and 4 were rounded to 6 digits.

Table 3
Finite-element solutions at $P_i = (0.5, \frac{1}{4}\sqrt{3})$ for Problem 5.1

Finite-element solution	$\bar{\xi}_i = 0.138933$	$\bar{\zeta}_i = 0.555732$
Exact solution	$u(P_i) = 0.138917$	$v(P_i) = 0.555667$

Table 4
Finite-element solutions at $P_i = (0.5, 0.5)$ for Problem 5.2

Finite-element solution	$\bar{\xi}_i = 0.666677$	$\bar{\zeta}_i = 0.333338$
Exact solution	$u(P_i) = 0.666667$	$v(P_i) = 0.333333$

References

[1] C.M. Ablow and C.L. Perry, Iterative solutions of the Dirichlet problem for $\Delta u = u^2$, *SIAM J. Appl. Math.* **7** (1959) 459–467.

[2] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* (Academic Press, New York, 1979).

[3] P.G. Ciarlet, Discrete maximum principle for finite difference operators, *Aequationes Math.* **4** (1970) 338–352.

[4] P.G. Ciarlet and P.A. Raviart, Maximum principle and uniform convergence for the finite element method, *Comput. Methods Appl. Mech. Engrg.* **2** (1973) 17–31.

[5] C.W. Cryer, The solution of quadratic programming problems using systematic overrelaxation, *SIAM J. Control* **9** (1971) 385–392.

[6] P.V. Dankwerts, *Gas Liquid Reactions* (McGraw-Hill, New York, 1970).

[7] K. Ishihara, On finite element schemes of the Dirichlet problem for a system of nonlinear elliptic equations, *Numer. Funct. Anal. Optim.* **3** (1981) 105–136.

[8] K. Ishihara, Explicit iterations with monotonicity for finite element approximations applied to a system of nonlinear elliptic equations, *J. Approx. Theory* **44** (1985) 241–252.

[9] K. Ishihara, Strong and weak discrete maximum principles for matrices associated with elliptic problems, *Linear Algebra Appl.* **88/89** (1987) 431–448.

[10] K. Ishihara, Successive overrelaxation method with projection for finite element solutions of nonlinear radiation cooling problems, *Computing* **38** (1987) 117–132.

[11] K. Ishihara, Successive overrelaxation method with projection for finite element solutions applied to the Dirichlet problem of the nonlinear elliptic equation $\Delta u = bu^2$, *Computing* **40** (1988) 51–65.

[12] C.S. Kahane, On a system of nonlinear parabolic equations arising in chemical engineering, *J. Math. Anal. Appl.* **53** (1976) 343–358.

[13] O.L. Mangasarian, Solution of symmetric linear complementarity problems by iterative methods, *J. Optim. Theory Appl.* **22** (1977) 465–485.

[14] O.L. Mangasarian and R. De Leone, Parallel successive overrelaxation methods for symmetric linear complementarity problems and linear programs, *J. Optim. Theory Appl.* **54** (1987) 437–446.

[15] G.T. McAllister, Difference methods for a nonlinear elliptic system of partial differential equations, *Quart. Appl. Math.* **23** (1966) 355–360.

[16] J.M. Ortega, *Numerical Analysis, A Second Course* (Academic Press, New York, 1972).

[17] J.M. Ortega and R.J. Plemmons, Extensions of the Ostrowski–Reich theorem for SOR iterations, *Linear Algebra Appl.* **28** (1979) 177–191.

[18] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970).

[19] J.M. Ortega and M.L. Rockoff, Nonlinear difference equations and Gauss–Seidel type iterative methods, *SIAM J. Numer. Anal.* **3** (1966) 497–513.

[20] J.S. Pang, On the convergence of a basic iterative method for the implicit complementarity problem, *J. Optim. Theory Appl.* **37** (1982) 149–162.

[21] J.S. Pang, Necessary and sufficient conditions for the convergence of iterative methods for the linear complementarity problem, *J. Optim. Theory Appl.* **42** (1984) 1–17.

[22] J.S. Pang and D. Chan, Iterative methods for variational and complementarity problems, *Math. Programming* **24** (1982) 284–313.

[23] M.H. Protter and H.F. Weinberger, *Maximum Principles in Differential Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1966).

[24] R.S. Varga, *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1962).

[25] D.M. Young, *Iterative Solution of Large Linear Systems* (Academic Press, New York, 1971).